DOCUMENT RESUME

ED 085 679 CS 000 877

AUTHOR	Ghatala, Elizabeth S.; Levin, Joel R.
TITLE	Developmental Differences in Frequency Judgments of Words and Pictures.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	TR-246
PUB DATE	Jan 73
CONTRACT	NE-C-00-3-0065
NOTE	21p.; Report from the Research Component, "Children's Learning and Development"
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Cognitive Processes; *Educational Research; Elementary Grades; *Learning; *Memory; Recall (Psychological); Verbal Ability; Verbal Development; *Verbal Learning; Visual Learning
IDENTIFIERS	*Verbal Discrimination

ABSTRACT

Children in kindergarten, third grade, and fifth grade were presented a list of either pictures or words (with items presented varying numbers of times on the study trail). In both picture and word conditions, half of the subjects estimated how many times each item had been presented (absolute judgments) and the other half judged which of two items had occurred more often on the study trail (relative judgments). The primary findings were: (1) ability to encode and discriminate frequency for both pictures and words increased with age, and (2) picture-word differences (in favor of pictures) increased with age. The results lend support to a two-factor model of developmental change in frequency judgment performance involving pictures and words. The model provides for predictions concerning performance on verbal discrimination and recognition memory tasks as a function of age and stimulus mode, and also suggests further research to explain picture-word differences in verbal learning tasks. (Author)

ED 085679

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Technical Report No. 246

DEVELOPMENTAL DIFFERENCES IN FREQUENCY JUDGMENTS OF WORDS AND PICTURES

by

Elizabeth S. Ghatala, Weber State College

and

Joel R. Levin, University of Wisconsin

Report from the Research Component
Children's Learning and Development

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin

January 1973

FILMED FROM BEST AVAILABLE COPY

05 000 877

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

Center Contract No. NE-C-00-3-0065

Statement of Focus

Individually Guided Education (IGE) is a new comprehensive system of elementary education. The following components of the IGE system are in varying stages of development and implementation: a new organization for instruction and related administrative arrangements; a model of instructional programing for the individual student; and curriculum components in prereading, reading, mathematics, motivation, and environmental education. The development of other curriculum components, of a system for managing instruction by computer, and of instructional strategies is needed to complete the system. Continuing programmatic research is required to provide a sound knowledge base for the components under development and for improved second generation components. Finally, systematic implementation is essential so that the products will function properly in the IGE schools.

The Center plans and carries out the research, development, and implementation components of its IGE program in this sequence: (1) identify the needs and delimit the component problem area; (2) assess the possible constraints—financial resources and availability of staff; (3) formulate general plans and specific procedures for solving the problems; (4) secure and allocate human and material resources to carry out the plans; (5) provide for effective communication among personnel and efficient management of activities and resources; and (6) evaluate the effectiveness of each activity and its contribution to the total program and correct any difficulties through feedback mechanisms and appropriate management techniques.

A self-renewing system of elementary education is projected in each participating elementary school, i.e., one which is less dependent on external sources for direction and is more responsive to the needs of the children attending each particular school. In the IGE schools, Center-developed and other curriculum products compatible with the Center's instructional programing model will lead to higher student achievement and self-direction in learning and in conduct and also to higher morale and job satisfaction among educational personnel. Each developmental product makes its unique contribution to IGE as it is implemented in the schools. The various research components add to the knowledge of Center practitioners, developers, and theorists.

Acknowledgments

We wish to thank the principal, Mr. William Smeaton, and the teachers and students of Edgerton Integrated Elementary School, Edgerton, Wisconsin, for their kind assistance in the conduct of this experiment. In addition, we are grateful to Nancy Hurlbut and Eve Inzer for collecting the data, to Ed Haertel for handling the computer programming, and to Billie Albrecht and Diane Eich for preparing the final draft of the paper. The order of authorship was determined randomly.

Table of Contents

		Page
	Acknowledgments	iv
	List of Tables and Figures	vii
	Abstract	ix
I.	Introduction	1
II.	Method	3
	Subjects	3
	Design and Materials	3
	Procedure	5
III.	Results	7
	Absolute Judgment Task	7
	Relative Judgment Task	9
IV.	Discussion	11
	References	15

List of Tables and Figures

Table		Page
1	Design of the Relative Frequency Judgment Task	4
2	Performance Medians on the Word and Picture Absolute Judgment Tasks for Subjects at Each Grade Level	8
3	Performance Medians on the Word and Picture Relative Judgment Tasks for Subjects at Each Grade Level	9

Figure		
1	Mean number of subjects in each grade responding correctly to words at the six positions in the relative judgment task.	10

Abstract

Children in kindergarten, third grade, and fifth grade were presented a list of either pictures or words (with items presented varying numbers of times on the study trial). In both picture and word conditions, half of the Ss estimated how many times each item had been presented (absolute judgments) and the other half judged which of two items had occurred more often on the study trial (relative judgments). The primary findings were: (a) ability to encode and discriminate frequency for both pictures and words increased with age, and (b) picture-word differences (in favor of pictures) increased with age. The results lend support to a two-factor model of developmental change in frequency judgment performance involving pictures and words. The model provides for predictions concerning performance on verbal discrimination and recognition memory tasks as a function of age and stimulus mode, and also suggests further research to explain picture-word differences in verbal learning tasks.

I
Introduction

The encoded situational frequency of events has been postulated to be one of the major discriminative attributes of memory (Underwood, 1969). There is evidence that discrimination of differences in situational frequency is the primary basis for performance in verbal discrimination tasks where S must learn to choose the correct member in each of a series of paired items (Ekstrand, Wallace, & Underwood, 1966), as well as in recognition memory tasks where S must discriminate between old and new items (Underwood & Freund, 1970b). Studies which have directly measured Ss' ability to judge frequency of events and to discriminate between events of differing presentation frequency have shown that frequency is well developed as a discriminative memory attribute in college students (Hintzman, 1969; Underwood & Freund, 1970a). One purpose of this study was to investigate age changes in the ability to encode and discriminate event frequency.

Children in kindergarten, third grade, and fifth grade were shown a series of items. Some items were shown only once, some were shown twice, and some three, four, or more times. In one type of task (absolute judgment task), Ss were presented with each item from the study trial and asked to judge how many times it had occurred. In the relative judgment task, Ss were asked to judge which of two items had occurred more often on the study trial. We expected that the ability to judge and discriminate event frequency as measured by these tasks would improve with age just as there is improvement in discrimination involving other attributes of memory, such as recency (Mathews & Fozard, 1970) and Osgood's evaluative dimension of word meaning (Cermak, Sagotsky, & Moshier, 1972).

In the experiment, absolute and relative frequency judgments were obtained for both word and picture stimuli. There is ample evidence indicating that pictorial materials are learned better and more rapidly than verbal materials in a variety of tasks including paired-associate learning (e.g., Levin, Davidson, Wolff, & Citron, in press; Rohwer, 1970), discrimination learning (e.g., Rowe & Paivio, 1971; Wilder & Levin, 1973), free recall learning (e.g., Paivio, Rogers, & Smythe, 1968; Sampson, 1970), and recognition memory (e.g., Jenkins, Neale, & Deno, 1967; Paivio & Csapo, 1969). Moreover, where evidence is available concerning age trends in the above tasks, there is an indication that the superiority of pictures over words increases with age. (For a discussion of the stimulus mode by age interaction in paired-associate learning, see Rohwer, 1970; in discrimination learning, see Wilder & Levin, 1973; and in free recall learning, see Levin, in press.)

The results of a recent study by Ghatala, Levin, and Wilder (1973) indicated that differences between pictures and words, at least in discrimination learning and recognition memory tasks, may be accounted for in terms of frequency theory as proposed by Ekstrand et al. (1966). In an absolute frequency judgment task (a) Ss judged pictures to be higher in situational frequency than words, even though the actual presentation frequencies of the two were the same; (b) Ss were less variable in their judgments of pictures than in their judgments of words; (c) Ss were able to identify "false alarms" (items that had not been previously presented) better when the materials were pictures than when they were words; and (d) though not reported by Ghatala et al., Ss' accuracy (in estimating exactly the presentation frequencies of items) was higher with pictures than words.

The Ghatala et al. findings, plus the assumptions of frequency theory concerning the manner in which frequency accrues to the right and wrong items in a discrimination learning task, clearly predict that the discrimination learning of pictures should be superior to the discrimination learning of words. Frequency theory also predicts (given the Ghatala et al. results) that performance on recognition memory

tasks should be better with pictorial than with verbal materials.

Ghatala et al. speculated that the differences between pictures and words in the absolute frequency judgment task might be due to differences in background (or preexperimental) frequency of pictures and words. That is, pictures or drawings used in experiments, while representing common, familiar objects, are in themselves relatively novel stimuli in that Ss are unlikely ever to have encountered the particular experimental pictures in the past. On the other hand, the words (or labels for the pictured objects) have probably been encountered by Ss many times in their past experience. Assuming the applicability of Weber's psychophysical law to frequency judgment tasks, one might posit that adding a frequency unit (experimentally) to items already high in frequency (preexperimentally), i.e., words, would be less noticeable than adding a frequency unit to items low in frequency, i.e., pictures. Carrying the above argument one step further--while incorporating the previous speculation that the ability to encode and discriminate event frequency increases with age--we were led to the following prediction: with increasing age, accuracy of frequency judgments and frequency discriminations involving pictures should become increasingly better than those involving words. It need only be assumed that the background frequency of words increases with age (or more precisely, with increased language experience) while the background frequency of pictures does not (since the particular pictorial representations are novel for Ss of all ages). Therefore, differences in background frequency between words and pictures should be greater for older children (high for words and low for pictures) than for younger children (low for both words and pictures). Accordingly, while the situational frequency of pictures was assumed to be easier to judge and discriminate at all age levels, based on the preceding rationale (and invoking Weber's Law) we expected to find increasing picture-word differences in frequency judgments and discriminations with increases in chronological age.

II
Method

Subjects

Forty-eight Ss from each of grades K, 3, and 5 participated in the experiment. The Ss attended an elementary school in a semirural Wisconsin community. The mean ages of the children were as follows--kindergarten: 5.7 years, with a range from 5.0 to 6.7 years; third grade: 9.2 years, with a range from 8.0 to 10.2 years; fifth grade: 10.9 years, with a range from 9.7 to 11.6 years.

Design and Materials

Half of the Ss in each grade were administered an absolute frequency judgment task, and half received a relative frequency judgment task. Within each task, half of the Ss were presented pictorial stimuli, and the other half were given verbal stimuli.

The pictorial stimuli consisted of 73 line drawings of common objects (e.g., a ball, a cat, a truck), while the verbal stimuli consisted of 73 names of similarly concrete objects.[1] Following selection of the total samples of pictures and words, the stimuli were randomly assigned to be used in either the absolute or relative judgment task.

Absolute Judgment Task

From the total samples of pictures and words, 49 of each were selected to be used in the picture and word conditions. In the picture condition, 39 of the pictures were randomly distributed among the four frequency categories represented in the study list. The ten remaining pictures were used as filler (or zero-frequency) items on the test trial. The study list consisted of 20 pictures presented once, ten pictures presented twice, six pictures presented three times, and three pictures presented four times, thereby making a total of 70 study presentations. Pictures with multiple occurrences were distributed equally in each approximately equal-sized section of the list, with the number of sections being determined by the frequency. That is, a picture presented twice occurred once in each half of the list, and so on. The same picture never occurred in adjacent positions. Of the 20 pictures occurring once, five were randomly assigned to each quarter of the list. The test list consisted of the 39 pictures presented for study plus the ten pictures which had not been presented. The order of the pictures on the test trial was random. The line drawings were photographed and mounted one to a slide transparency. Thus, there were 70 slides presented for study

---

[1] The present data were obtained within the context of a larger study, the purpose of which was to examine intertask correlations. That is, each S received two absolute judgment tasks (one with pictures and one with words) and two relative judgment tasks (a picture and a word task), with the order of the four tasks counterbalanced across Ss. The data reported here were taken only from the first stage of practice in order to provide the purest test of the present hypotheses. For obvious reasons the design of the larger study precluded the use of words which were the labels of the pictures. However, all words were selected from a homogeneous, high-frequency (Thorndike & Lorge, 1944) pool, and were uniformly high in meaningfulness and concreteness according to the Paivio, Yuille, and Madigan (1968) norms. Thus, while the words were not identical to the pictures' labels, within the limits of random sampling they would be expected to be equivalent in meaningfulness, concreteness, and the like.

followed by 49 test slides.

In the word condition, words were randomly assigned in the same numbers to the same four frequency categories as in the picture condition. The order of occurrence of items representing different frequency categories on both presentation and test trials was exactly the same in the two conditions. The words were presented auditorily to all Ss because of obvious differences in reading ability among the three age groups.[2]

Relative Judgment Task

The remaining 24 pictures were used in the relative judgment task. To provide direct evidence concerning the operation of Weber's Law when frequency is manipulated in the laboratory, the items paired during testing represented a unique combination of base frequency (the frequency of the less frequent member of the pair) and absolute frequency difference, as shown in Table 1. Four base frequencies (0, 1, 3, and 5) and three absolute frequency differences (1, 2, and 3) for each base frequency accounted for the 12 test pairs. The major predictions stemming from Weber's Law are that while accuracy of discrimination should increase with increasing absolute frequency difference, increasing the base frequency of pairs given the same absolute frequency difference should decrease the accuracy of discrimination.

To achieve the necessary induced frequencies required 78 presentations: four pictures were presented once; two, twice; five, three times; two, four times; four, five times; two, six times; one, seven times; and one, eight times. In addition, three pictures representing the base frequency of zero were not presented for study, appearing for the first time on the test trial. All pictures were randomly assigned to the above frequencies.

The ordering of the pictures across the 78 positions was random, subject to the restriction that those with multiple occurrences appear equally often in each approximately equal-sized section of the list, with the number of sections being determined by the frequency. The same picture never occurred in adjacent positions. The four pictures that occurred once were assigned randomly to each of four sections. The line drawings were

TABLE 1
DESIGN OF THE RELATIVE FREQUENCY JUDGMENT TASK

Base Frequency	Difference Frequency		
	1	2	3
0	0 vs. 1[a] (III)	0 vs. 2 (II)	0 vs. 3 (I)
1	1 vs. 2 (IV)	1 vs. 3 (III)	1 vs. 4 (II)
3	3 vs. 4 (V)	3 vs. 5 (IV)	3 vs. 6 (III)
5	5 vs. 6 (VI)	5 vs. 7 (V)	5 vs. 8 (IV)

[a] Cell entries represent the situational frequencies of the items being compared. Roman numerals will be described in a later section.

[2] In this regard, it should be noted that words were presented auditorily in this experiment solely to permit picture-word inferences for Ss as young as kindergartners. At the same time, this represents a departure from the Ghatala et al. (1973) experiment, where the words were displayed visually with no auditory accompaniment. The differing presentation modes of the verbal stimuli will be attended to, therefore, when results of the two experiments are compared.

photographed and mounted one to a slide for presentation. On the test trial the pictures appeared side by side on a single slide. The low-frequency member of the pairs occurred equally often in the left- and right-hand positions across the 12 test slides. The order of the test pairs was random.

In the word condition, words were randomly assigned in the same numbers to the same frequency categories as in the picture condition. On both presentation and test trials the order of occurrence of items representing different frequency categories was the same in the two conditions. As was the case for the absolute judgment task, all words were presented auditorily. On the test trial, the low-frequency member of each pair was spoken first or second an equal number of times, corresponding to the left-right positions of the low-frequency members of the picture pairs.

Procedure

Subjects were tested individually in a small private room within the school building. Preceding each frequency judgment task, a short practice trial (utilizing items not used in the regular list) was given to insure that S understood the nature of his task.

In the absolute judgment task with pictures, the Ss were told that they would see pictures of common objects and that many of the pictures would be shown more than once. They were instructed that after they had seen all the pictures, they would be asked to judge how many times they had seen each picture on the study trial. Performance on an eight-item practice list (three pictures presented once; one picture, twice; one picture, three times) indicated that Ss in all grades understood the task, and that the kindergarten Ss could properly assign numbers to judge event frequency. On the study trial the pictures were presented at a 2-second rate. The test trial was essentially unpaced, but most Ss responded within 2 seconds. Subjects were required to respond to each item on the test trial, guessing if uncertain.

In the relative judgment task with pictures, Ss were told that they would see pictures, some of which would occur a few times and some, many times. They were instructed that following the presentation of the pictures they would have to pick out the pictures they had seen the most often. An eight-item practice list was presented followed by three test pairs (frequencies: 1-3, 1-2, 1-0). The presentation rate was 2 seconds. The test trial was unpaced but, again, most Ss responded within 2 seconds. The Ss were required to choose the more frequent member of each pair, guessing if uncertain. Subjects were allowed to indicate their choice in any manner (e.g., naming the object, pointing) as long as their choice could be clearly interpreted by E. All pictures in both the relative and absolute judgment tasks were presented by means of a Kodak Carousel slide projector.

The procedures for the absolute and relative judgment tasks employing words paralleled those for pictures. The words were read by E at a 2-second rate. The E was well practiced on both word lists and attempted to read the words with equal emphasis and consistent inflection. The rate of reading was regularly checked by means of a stop watch. Test trial items were also read by E. On the relative judgment task, Ss were allowed to indicate their choice by a variety of means (e.g., saying the word, saying "first" or "second") as long as the choice could be interpreted by E.

III
Results

Performance on the absolute and relative frequency judgment tasks is summarized in Tables 2 and 3. As a result of two outliers in the absolute judgment task for words in the kindergarten sample, marked skewness was detected. Consequently, average performance is reported in terms of medians. At the same time, the large differences in between-S variability from grade to grade indicated that nonparametric rank tests would provide for more efficient analyses of the data.

In order to test the hypothesis that the ability to encode and discriminate situational frequency improves with age, separate analyses were performed on the word and picture tasks. Further, a joint analysis in corporating both word and picture data was conducted to test the hypothesis that picture-word differences in frequency judgment ability increase with age.

Absolute Judgment Task

Based on the Ghatala et al. (1973) experiment, mean and variability measures reflecting absolute judgments of words and pictures were computed. For each S, the mean of all frequency judgments for items with the same presentation frequency was determined (i.e., a mean on the "one" items, a mean on the "two" items, etc.). A within-S variance measure, based on S's mean squared deviation for each presentation frequency, represents the degree to which S consistently judged items of the same presentation frequency as being the same.

In addition, the number of correct test list identifications of actual presentation frequency (out of 49) was computed for each S as an index of his "accuracy" in judging situational frequency. Although this measure was not used by Ghatala et al. (1973) it was noted subsequently that such an index tends to be highly correlated with, but perhaps more sensitive than, the within-S variability measure in absolute frequency judgment tasks (Wilder, Levin, Ghatala, & McNabb, in press).

Although multiple correlated measures of performance were calculated (see Table 2), multivariate analyses were not performed since it was desired to identify variable-by-variable effects (Levin & Marascuilo, 1972). At the same time, a priori significance levels could not be easily justified. Thus, the significance probability corresponding to each dependent variable is reported for the reader's information.[3]

In each analysis, a contrast involving age was formed. Since all hypotheses were explicit in nature, linear trend coefficients were fitted to the unequally spaced K, 3, 5 factor in order to determine whether monotonic changes occurred from grade to grade. (These coefficients turned out to be $-2\frac{2}{3}$, $+\frac{1}{3}$, and $+2\frac{1}{3}$, respectively.) Similarly, since all hypotheses were explicit with regard to their directionality, all significance probabilities were computed as one-tailed values.

To test age hypotheses for each dependent variable, the data were rank-ordered (separately for pictures and words), as would be done for a Kruskal-Wallis analysis of variance. Following this, a linear contrast involving grades was formed on the mean ranks. The contrast was then tested for statistical significance using a large-sample Z approximation (Marascuilo & McSweeney, 1967), a summary of which is provided in the last two columns of Table 2.

In the first four rows of Table 2 are the medians of Ss' mean frequency judgments, by grade, for items with presentation frequencies of one, two, three, and four (the latter three having been averaged). The next four rows represent the median within-S variances corresponding to the same items, and in the last

---

[3] Since a correction for tied ranks was not employed, all significance probabilities will be on the conservative side.

TABLE 2
PERFORMANCE MEDIANS ON THE WORD (W) AND PICTURE (P) ABSOLUTE JUDGMENT TASKS FOR SUBJECTS AT EACH GRADE LEVEL

Measure	Grade			Contrast	
	K	3	5	Z	p <
1 Mean (W)	1.02	1.10	1.00	Wrong direction	
1 Mean (P)	1.25	.98	1.02	Wrong direction	
2-4 Mean (W)	1.78	2.50	2.32	1.35	.09
2-4 Mean (P)	2.84	2.76	3.00	.45	.33
1 Variance (W)	.58	.35	.22	2.82	.003
1 Variance (P)	.29	.22	.16	1.86	.04
2-4 Variance (W)	.74	.48	.48	1.71	.05
2-4 Variance (P)	1.17	.38	.20	3.50	.001
Total Correct (W)	20.5	23	27	2.66	.004
Total Correct (P)	26.5	36.5	36	3.70	.001

row the median number of correct frequency identifications is reported. Performance on the "zero" items is not presented in Table 2, but will be mentioned where appropriate.

Age Differences

It had been predicted that frequency judgment ability would improve with age. Based on the statistical analysis of the linear contrasts for the ranked data, this was found to be the case for the word and picture variances, which exhibited a systematic decrease (all p's < .05), as well as for total correct, which showed an increase (both p's < .004). On the other hand, mean frequency judgments did not increase reliably with age, and, in fact, two of these contrasts were in the wrong direction. [A moment's reflection will suggest that these latter measures (i.e., mean judgments) do not really represent frequency judgment ability (i.e., the ability to estimate correctly event frequency), and thus the data associated with them are not inconsistent with the predictions as stated.] Concerning the age trend on the "zero" items, 38% of the 24 kindergarten Ss got all ten identifications correct, while this figure was 71% and 67% for third and fifth graders, respectively.

Picture-Word Differences by Age

In order to analyze the data bearing on the hypothesis of major interest in the present study, viz. that picture-word differences would increase with age, the following strategy was adopted for each dependent measure. Before the data were ranked, each S's score was "aligned" on the median score for S's grade. Operationally, this consisted of subtracting the median score for a particular grade (across word and picture tasks) from all scores in that grade. Conceptually, this is analogous to removing the grade main effect from the data which permits a more sensitive test of treatment differences. (Note that in doing this, picture-word differences within grades will not change, but scores on both pictures and words become more or less "standardized" from grade to grade.) Following alignment, the data were ranked across the six samples formed by the combination of three age groups (K, 3, 5) and two stimulus modes (words, pictures), and then analyzed according to an extension of the procedures proposed by Hodges and Lehmann (1962). Essentially, a contrast involving the picture and word mean ranks in each grade was formed, with the previously described linear trend coefficients applied to the picture-word differences in mean rank

(actually a nonparametric test for a linear interaction). A Z statistic, based on a large sample approximation (McSweeney & Marascuilo, 1970) was then computed.

The results of this analysis are now described, with the P-W differences in Table 2 providing a helpful point of reference. No significant increase in picture-word differences was detected for mean frequency judgments, either on items presented once (Z was in the wrong direction) or on those presented two to four times ($Z = .06$, $p < .48$). Neither was there a significant increase in the differences between word and picture variances on the 1 items (Z was in the wrong direction).

However, on the 2-4 items, the variance of words relative to that of pictures did increase from K through 5 ($Z = -3.00$, $p < .002$). The sensitive "total correct" measure also reflects this change, with Ss' differential ability to identify correctly situational frequencies of pictures and words exhibiting an increase (in favor of pictures over words) with age ($Z = 2.02$, $p < .03$).

No important differences were noted on the "zero" items, where the difference in percentages of Ss who identified all ten items correctly was 25% (in favor of pictures) for kindergartners, and 25% and 33% (in the same direction) for third and fifth graders, respectively.

Picture-Word Differences

Though not assessed statistically in the present study (since it was not of primary interest), a descriptive comparison of picture and word judgments may be obtained from Table 2. In five out of six comparisons, pictures had higher mean judgments associated with them than words; while in another five out of six comparisons, pictures exhibited a smaller variance than words. These results substantiate the previously discussed Ghatala et al. (1973) findings. Note also that in each grade, more pictures than words had correctly identified frequencies. Finally, judgments of the ten "zero" items resulted in perfect performance for 26 out of 36 (72%) picture Ss, but for only 16 out of 36 (44%) word Ss.

Relative Judgment Task

The same statistical treatment of the data that was employed for the absolute judgment task was used here. As may be seen from the median performance on the 12-item relative judgment task (Table 3), discrimination of event frequency was quite good at all three grade levels. As was true previously, the median performance on the picture task was descriptively superior to that on the word task within each grade.

TABLE 3
PERFORMANCE MEDIANS ON THE WORD AND PICTURE RELATIVE JUDGMENT TASKS FOR SUBJECTS AT EACH GRADE LEVEL

	Grade			Contrast	
	K	3	5	Z	p <
Word	9.0	10.0	9.5	1.66	.05
Picture	9.5	11.0	12.0	3.21	.001

Concerning the two hypotheses of the present study, both were supported in that (a) the word and picture tasks each revealed an increase in correct discriminations with age (p's < .05 and .001 respectively); as well as (b) the picture-word difference showed a significant ($Z = 2.63$, $p < .005$) increase with age.

Thus, it may be safely concluded that the results of the relative frequency judgment task parallel those of the "total correct" measure on the absolute judgment task. That is, both are sensitive to the predicted age changes. This finding is not surprising if it is interpreted as reflecting two related components of frequency judgment ability: "estimation accuracy" (as measured by total correct in the absolute judgment task) and "discrimination" of event frequency (represented by number correct on the relative judgment task). In both cases, the ability to judge frequency improves with age, with the superiority of picture judgments over word judgments exhibiting a similar increase.

Weber's Law

Recall that the relative judgment task consisted of 12 items, each representing a different base frequency/difference frequency combination (see Table 1). Frequency theory (Ekstrand et al., 1966) predicts that discriminations should be easier with larger differences in presentation frequency when base frequency is held constant (within rows, moving from left to right). At the same time, Weber's Law predicts that discriminations should be more difficult with larger base frequency when the difference frequency

is held constant (within columns, moving from top to bottom). Therefore, Weber's Law within the context of frequency theory leads to the general prediction that the easiest discriminations should occur for items in the upper right cells of Table 1, and the most difficult, for items in the lower left. Accordingly, the 12 cells were partitioned into six different groups or "positions" (indicated by the Roman numerals in Table 1) to reflect the general prediction. The hypothesis tested was based on Ss' performance at each position and may be stated simply as I > II > III > IV > V > VI.

To test this hypothesis, S's performance on the 12-item task was separated by position and, where more than one cell was involved, averaged. A linear contrast defining a monotonic decrease across positions was then formed from the six position scores for each S. Initially, this procedure was to be followed for both word and picture tasks. However, since performance on the picture task was so high, especially for third- and fifth-grade Ss (see Table 3), a reasonable test of Weber's Law could not be made with these data. Consequently only the word task data were analyzed.

Although the results for all Ss given the word task are presented in Figure 1, the scores of four Ss (one in kindergarten, two in third grade, and one in fifth grade) were not included in the statistical analysis since they got all 12 items correct, i.e., they hit the task's ceiling, which would not provide for a valid test of the hypothesis.

Nonparametric rank tests were used to determine whether the predicted ordering was obtained within each grade, as well as to determine whether the effect changed from grade to grade. Note that according to previously discussed rationales, given the same experimentally induced base and difference frequencies, Weber's Law should be less manifest for older than for younger Ss (with word stimuli). This is because the background frequencies of such items for older Ss are already high; consequently, their discriminations should be less affected by experimental manipulations of frequency.

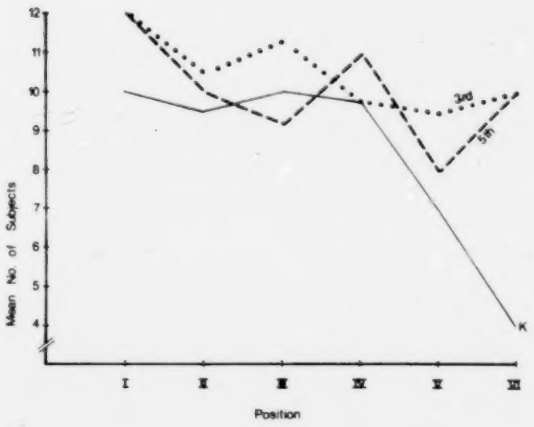

Figure 1. Mean number of subjects in each grade responding correctly to words at the six positions (based on Weber's Law) in the relative judgment task.

Concerning the operation of Weber's Law within each grade, the Wilcoxon signed rank test was used to decide whether the distribution of scores which conformed to the pattern (i.e., positive contrasts) differed from what would be expected by chance. Based on acceptable significance probabilities, it was concluded that Weber's Law was supported in the kindergarten ($p < .007$) and third-grade ($p < .01$) samples, but not in the fifth-grade sample ($p < .14$). As corroborative evidence for this, a nonparametric test of the contrast associated with the Weber's Law by grade interaction confirmed that the operation of the law did indeed diminish with age ($Z = 1.84$, $p < .04$), as the flatter third- and fifth-grade curves in Figure 1 would suggest.

IV
Discussion

The results of the experiment clearly support the prediction that Ss' ability to encode and discriminate event frequency improves with age. Measures such as within-S variance and total correct identifications in the absolute judgment task, which reflect Ss' ability to estimate event frequency consistently and accurately, indicated an improvement with age for both words and pictures. The measure of Ss' ability to discriminate event frequencies (i.e., number correct in the relative judgment task) also showed that there is improvement in this ability with age for both words and pictures.

The second major prediction, that of increasing superiority with age for judgments involving pictures as compared to words, was also supported by the data. This interaction of age with stimulus mode was found for one of the measures of within-S variance (for presentation frequencies of two to four) and for total number of correct identifications in the absolute judgment task, and for number correct in the relative judgment task. As with the main effect of age, the stimulus mode by age interaction was found for those measures reflecting Ss' ability to estimate and discriminate event frequency.

Scrutiny of the age changes in the above measures for pictures and for words suggests that two independent factors may well be contributing to the stimulus mode by age interaction. Note that the general form of this interaction (see Tables 2 and 3) reveals that although frequency judgments of both pictures and words improve with age, those based on pictures improve at a faster rate. The increases for both pictures and words may be attributable to a frequency judgment ability factor, which improves with age for either type of material. The fact that improvement of judgments involving words was not as great as that for pictures reflects the differential operation of a second factor which we earlier referred to as background frequency.

We have assumed that with increasing age an increasing amount of background or pre-experimental frequency accrues to words, while there is little or no accrual of background frequency to the experimental pictures with age. The assumption that accurate frequency estimations and discriminations are more difficult for items of higher background frequency (Weber's Law) is useful in accounting for the stimulus mode by age interaction in a straightforward manner. For pictures, the only factor operating is the general improvement with age in the ability to encode and discriminate situational frequency. For words, on the other hand, both a positive factor (improving ability) and a negative factor (increasing background frequency) are affecting performance at each age. Thus, Ss' net performance at a particular age may be regarded as the sum of two counteracting influences, each assumed to increase with age. (Note that for the present data, since there is improvement with age on the word frequency judgment tasks, the first factor would be hypothesized to be the more potent of the two.)

It has been demonstrated in the present study and by others (Hintzman, 1969; Underwood & Freund, 1970a) that predictions stemming from Weber's Law hold when frequency is manipulated in the laboratory. That is, as base frequency increases (given a constant absolute difference frequency) errors in discrimination increase. Moreover, the present result of increasing picture-word differences with age is consistent with the notion that frequency established outside the laboratory also follows the same general law. Thus, Weber's Law can be invoked to describe either the behavior of Ss in a relative judgment task where base frequencies are manipulated experimentally, or the behavior of Ss of different ages in judging the frequency of pictures and words. In the latter case the "base frequencies" of words are assumed to increase with age (or increased language experience) while the "base frequencies" of pictures of the type used in this and

most other learning experiments are assumed to remain relatively constant (and low) with increasing age due to Ss' relative lack of exposure to these stimuli at all ages. Following the above reasoning, we were able to predict the stimulus mode by age interaction obtained in the present experiment.

The implications of the present results for verbal discrimination and recognition memory tasks should be obvious. Since performance in both tasks is postulated to be based on discrimination of frequency differences between correct and incorrect items, we would predict increasing superiority of pictures over words in both tasks as a function of age. There are data from the present study which speak directly to this issue.

Consider first the verbal discrimination task where S must learn to choose the "correct" item in each of a series of paired items. According to frequency theory (Ekstrand et al., 1966), on the first anticipation trial of such a task S is assumed to make a representational response (RR) to each item (adding one frequency unit to each), a pronunciation response (PR) to one of the items (adding one frequency unit to the chosen item), and, following feedback, S may make a rehearsal of the correct response (RCR) to the correct item. For those pairs in which S's initial guess was correct, a three to one frequency difference exists favoring the correct item. Since there is no reason to assume that the probability of guessing the correct member of a pair on the first trial in verbal discrimination learning varies as a function of stimulus modes, an equal number of pairs should have the same three to one difference in a list consisting of word pairs as in one consisting of picture pairs. Thus, S's ability to discriminate this frequency difference for word and for picture pairs is the critical factor influencing rate of learning.

One of the items in the word and picture relative judgment tasks involved a three to one frequency discrimination (see Table 1). The percentages of Ss (out of 12) who correctly chose the more frequent item in the word pair were 83%, 92%, and 75% in kindergarten, third grade, and fifth grade, respectively. For the picture pair, the corresponding percentages for each grade were 75%, 92%, and 100%. Thus, descriptively the difference between pictures and words in ease of discriminating a three to one frequency difference increased with age (-.08, .00, and .25, respectively). In fact, at the youngest age, word pairs appeared to be more easily discriminated than picture pairs. While we would not want to argue on the basis of these limited data that such a crossover effect for pictures and words would occur as a function of age in discrimination learning, increasing picture-word differences with age might be expected. This type of interaction of stimulus mode with age has been suggested in at least one discrimination learning experiment (Wilder & Levin, 1973).

In the relative frequency judgment task, Ss were also tested on items involving a zero to one frequency discrimination (see Table 1). This kind of item corresponds to a recognition memory task where S is required to choose between items he has seen on the study trial and items he has not seen. Again, the percentage of Ss correctly choosing the more frequent item in the relative judgment task varied as a function of stimulus mode and age. For the word stimuli, 83% of the kindergartners, 100% of the third graders, and 67% of the fifth graders correctly discriminated the more frequent member of the pair. In the picture condition the corresponding percentages at each grade level were 92%, 100%, and 100%. Again, descriptively the difference favoring pictures increases with age (.09, .00, and .33, respectively). Based on this finding, we would anticipate an interaction of stimulus mode and age in recognition memory. To our knowledge, such an effect has yet to be investigated.

Based on the two-factor model which has been proposed as an explanation for stimulus mode by age interactions in certain kinds of learning tasks, additional hypotheses may be derived. For example, we posited earlier that the background frequencies of auditorily presented words were low for the kindergartners of the present study. As children learn to read, however, one might expect to find differences in background frequency between words presented auditorily and words presented visually. That is to say, while the background frequencies of words might still be presumed low for first or second graders (novice readers), one might expect them to be relatively lower for only recently encountered stimuli ("printed" words) than for more familiar ones ("listened to" words). These notions, which when adapted to the stimulus mode by age interaction hypothesis would predict differential effects for visually and auditorily presented words, are among those which we are currently investigating.

The present results, in conjunction with those of Ghatala et al. (1973), further implicate "frequency" as the primary attribute in discrimination learning (Ekstrand et al., 1966; Underwood, 1969) particularly with regard to accounting for differences in the ease of learning pictures and words. In view of the data from a recent experiment wherein picture-word differences in discrimination learning were obliterated when the two types of material

GPO 803-302-3

were equated in terms of independently obtained frequency judgments (Levin, Ghatala, & Wilder, in press), explanations couched in frequency theory may not be as far-fetched as one might think.

References

Cermak, L. S., Sagotsky, G., & Moshier, C. Development of the ability to encode within evaluative dimensions. Journal of Experimental Child Psychology, 1972, 13, 210-219.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. Psychological Review, 1966, 73, 556-578.

Ghatala, E. S., Levin, J. R., & Wilder, L. Apparent frequency of words and pictures as a function of pronunciation and imagery. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 85-90.

Hintzman, D. L. Apparent frequency as a function of frequency and the spacing of repetition. Journal of Experimental Psychology, 1969, 80, 139-145.

Hodges, J. L., & Lehmann, E. L. Rank methods for combination of independent experiments in analysis of variance. The Annals of Mathematical Statistics, 1962, 33, 482-497.

Jenkins, J. R., Neale, D. C., & Deno, S. L. Differential memory for picture and word stimuli. Journal of Educational Psychology, 1967, 58, 303-307.

Levin, J. R. Verbal and visual processes in children's learning: IV. An attempt to reduce picture-word differences in free recall learning through concretization. Wisconsin Research and Development Center for Cognitive Learning, Working Paper No. 106, in press.

Levin, J. R., Davidson, R. E., Wolff, P., & Citron, M. A comparison of induced imagery and sentence strategies in children's paired-associate learning. Journal of Educational Psychology, in press.

Levin, J. R., Ghatala, E. S., & Wilder, L. Picture-word differences in discrimination learning: I. Apparent frequency manipulations. Wisconsin Research and Development Center for Cognitive Learning, Techical Report No. 253, in press.

Levin, J. R., & Marascuilo, L. A. Type IV errors and interactions. Psychological Bulletin, 1972, 78, 368-374.

Marascuilo, L. A., & McSweeney, M. Nonparametric post hoc comparisons for trend. Psychological Bulletin, 1967, 67, 401-412.

Mathews, M. E., & Fozard, J. L. Age differences in judgments of recency for short sequences of pictures. Developmental Psychology, 1970, 3, 208-217.

McSweeney, M., & Marascuilo, L. A. Aligned rank statistics for the randomized block design. Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, 1970.

Paivio, A., & Csapo, K. Concrete-image and verbal memory codes. Journal of Experimental Psychology, 1969, 80, 279-285.

Paivio, A., Rogers, T. B., & Smythe, P. C. Why are pictures easier to recall than words? Psychonomic Science, 1968, 11, 137-138.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76 (No. 1, Part 2).

Rohwer, W. D., Jr. Images and pictures in children's learning: Research results and instructional implications. In H. W. Reese (Chm.), Imagery in children's learning: A symposium, Psychological Bulletin, 1970, 73, 393-403.

Rowe, E. J., & Paivio, A. Discrimination learning of pictures and words. Psychonomic Science, 1971, 22, 87-88.

Sampson, J. R. Free recall of verbal and nonverbal stimuli. Quarterly Journal of Experimental Psychology, 1970, 22, 215-221.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College Bureau of Publications, 1944.

Underwood, B. J. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Underwood, B. J., & Freund, J. S. Relative frequency judgments and verbal-discrimination learning. Journal of Experimental Psychology, 1970, 83, 279-285. (a)

Underwood, B. J., & Freund, J. S Testing effects in recognition of words. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 117-125. (b)

Wilder, L., & Levin, J. R. A developmental study of pronouncing responses in the discrimination learning of words and pictures. Journal of Experimental Child Psychology, 1973, 15, 278-286.

Wilder, L., Levin, J. R., Ghatala, E. S., & McNabb, S. Pronunciation and apparent frequency in a between-subjects design. Journal of Experimental Psychology, in press.

GPO 803-392-2